UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/980,373	11/29/2001	Gianluca Francini	22082	8445

7590 12/10/2003

Herbert Dubno
The Firm of Karl F Ross
5676 Riverdale Ave.
Box 900
Riverdale (Bronx), NY 10471

EXAMINER	
PAPPAS, PETER	
ART UNIT	PAPER NUMBER
2671	5

DATE MAILED: 12/10/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/980,373	FRANCINI ET AL.	
	Examiner	Art Unit	
	Peter-Anthony Pappas	2671	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *29 November 2001*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-7* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-7* is/are rejected.

7)☒ Claim(s) *1* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *11/29/2001* is/are: a)☐ accepted or b)☒ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. §§ 119 and 120

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☒ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

13)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application) since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.

a) ☐ The translation of the foreign language provisional application has been received.

14)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121 since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449) Paper No(s) *2*.
4) ☐ Interview Summary (PTO-413) Paper No(s). _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: .

DETAILED ACTION

Specification

1. The disclosure is objected to because of the following informalities: the version of the referenced ISO/IEC 14496 standard is not recited.

2. The specification is objected to as failing to provide proper antecedent basis for the claimed subject matter. See 37 CFR 1.75(d)(1) and MPEP § 608.01(o). Correction of the following is required: the term viseme is given a meaning repugnant to the usual meaning of the term when it is stated "...does not therefor refer to high-level MPEG-4 parameters."

3. Appropriate correction is required.

Drawings

4. The drawings Fig. 1, 3-4 and 6, are objected to because they are presented in a unclear manner and are illegible – the print is generally faded and the images are too dark to substantially make out any details. A proposed drawing correction or corrected drawings are required in reply to the Office action to avoid abandonment of the application. The objection to the drawings will not be held in abeyance.

5. Applicant is required to submit a proposed drawing correction in reply to this Office action. However, formal correction of the noted defect may be deferred until after the examiner has considered the proposed drawing correction. Failure to timely submit the proposed drawing correction will result in the abandonment of the application.

Claim Objections

6. Claim 1 is objected to because of the following informality: claims must be limited to a single sentence. Appropriate correction is required.

Claim Rejections - 35 USC § 112

7. The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

8. Claims 1-7 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

9. In regards to claims 1-7: the claims are generally narrative and indefinite, failing to conform with current U.S. practice.

10. In further regard to claim 1: the version of the referenced ISO/IEC 14496 standard is not recited.

11. In further regards to claim 2: the variables $ß_n(t)$ and t_n are not recited. In addition the equations disclosed are unclear as to how $ß_n(t)$ can be assigned result from three different functions when each of the three function shares at least one included limit point, within their respective closed intervals, with that of another function. For example $t \in [t_{n-1}, t_{n+1}]$, which yields $ß_n(t) = 0$, shares closed interval limit points of $t \in [t_n, t_{n+1}]$ and $t \in [t_{n-1}, t_n]$.

12. In further regards to claim 5: it is unclear as which elements of claim 5 pertain to "..., as described in the standard ISO/IEC 14496." Thus, it is understood this is applied to "...low-level facial animation parameters,..."

Claim Rejections - 35 USC § 102

13. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

14. Claims 1, 3 and 4 are rejected under 35 U.S.C. 102(b) as being unpatentable over Ostermann (Animation of Synthetic Faces in MPEG-4).

15. In regards to claim 1:

(a) an analytic phase, in which an alphabet of visemes is determined, i.e. a set of information representing the shape of a face of a speaker corresponding to phonetic units extracted from a set of audio training signals, and a synthesis phase, in which he audio driving signal is converted into a sequence of phonetic units associated to respective temporal information, whereas the sequence of visemes, corresponding to the phonetic units of the set comprised in the audio driving signal, are determined in the analytic phase, and the transforms required to reproduce the sequence of the visemes are applied to the model

16. Ostermann discloses:

- MPEG-4 defines a user terminal that allows decoding, composing and presenting multiple audio-visual objects. These A/V objects can be music, speech, synthesized speech from text-to-speech (TTS) synthesizer, synthetic audio, video sequences, arbitrarily shaped moving video objects, images, 3D computer animated models or synthetic face models. See page 6, column 1 paragraph 1.

- For synthetic visual contents, MPEG-4 allows to build 2D and 3D objects composed of primitives like rectangles, sphere, indexed face sets and arbitrarily shaped 2D objects. Special 3D objects are human faces and bodies. See page 1, column 1 paragraph 3 and column2 paragraph 1.
- The second output interface of the synthesizer send the phonemes of the synthesized speech as well as start time and duration information for each phoneme to the Phoneme/Bookmark-to-FAP-Converter. FAP is a face animation parameter. The converter translates the phonemes and timing information info face animation parameters that the face renderer uses in order to animate the face model. The precise method of how the converter derives visemes from phonemes is not specified by MPEG and left to the implementation of the decoder. In the current MPEG4 standard, the encoder is expect to send a FAP stream containing FAP number and value for every frame, to enable the receiver to produce desired facial actions. See page 6, column 2 paragraph 1.
- The FAP set contains the parameters visemes and expressions. See page 2, column 2 paragraph 1, and Table 1.

(b) said analytic phase provides an alphabet of visemes, determined as active shape model parameter vectors, to which the respective transforms of the model, expressed as parameters of low-level facial animation compliant with standard ISO/IEC 14496, are associated.

17. Ostermann discloses:

- The *FaceDefTable* specifies the type of transformation and a neutral value for the chosen transformation. During animation, the received value for the FAP and the neutral value determine the actual value. See page 3, column 2 paragraph 2.
- For the head, more then 70 model-independent animation parameters defining low-level actions are standardized. See page 1, column 1 paragraph 1.
- FAPs represent a complete set of basic facial actions including head motion, tongue, eye and mouth control. They allow the representation of natural facial expressions. They can also be used to define facial action units. See page 2, column 1 paragraph 5 and Table 1.

(c) During both the analytic phase and the synthesis phase, the sequences of visemes, corresponding to the phonetic unit, respectively, are transformed into continuous representation of movement by means of viseme interpolation, conducted as convex combinations of the visemes themselves to which combination coefficients, which are continue function of time, are associated, the combination coefficients carious in the synthesis phase being the same as those used for the analytic phase combination.

18. Ostermann discloses:

- In order to allow for coarticulation of speech and mouth movement transitions from one viseme to the next are defined by blending the two visemes with a weighting factor. See page 2, column 2 paragraph 1.

- For coding of facial animation parameters, MPEG-4 provides two tools. Coding of quantized and temporally predicted FAPs using an arithmetic coder allow for coding of FAPs and introducing small delay only. Using a discrete cosine transform (DCT) for coding a sequence of FAPs introduces significant delay but achieves higher coding efficiency. See page 4, column 2 paragraph 1.
- The first set of FAP values $FAP(i)_0$ at time instant 0 is coded in intra mode. The value of an FAP at time instant $k*FAP(i)_k$ is predicted using the previously decoded value $FAP(i)_{k-1}$. See page 4, column 2 paragraph 2.

19. It is noted that the use of the mathematical terms weighting factor and coefficient are considered interchangeable.

20. In regards to claim 3:

Method according to claim 1, characterized by the fact that the wire-frame vertices, corresponding to the model feature points, on the basis of which facial animation parameters are determined in the analytic phase, are identified and said viseme interpolation operations are conducted by applying transformations on feature points for each viseme, for animating a wire-frame based model.

21. Ostermann discloses:

- An *IndexedFaceSet* node defines the geometry (3D mesh) and surface attributes (color, texture) of a polygonal object. Texture maps are coded the wavelet coder of the MPEG texture coder. Since the face model is specified with a scene graph,

this face model can be easily extended to a head and shoulder model. See page 3, column 2 paragraph 1.

- A *FaceDefTable* defines how a model is deformed as a function of the amplitude of the FAP. It specifies, for a FAP, which Transform nodes and which vertices of an IndexedFaceSet node are animated by it and how. FaceDefTables are considered to be part of the face model. See page 3, column 2 paragraph 2.
- In order to define face animation parameters for arbitrary face models, MPEG-4 specific 84 feature points located in a face in order to provide a reference for defining facial animation parameters. See page 2, column 1 paragraph 4.
- Feature points may be used to define the shape of a proprietary face model. FAPs are defined by the motion of feature points. See page 3, Figure 2.

22. It is noted that the applicant discloses that a three-dimensional mesh structure is also referred to as a wire-frame. Thus, the expectations of the claim are still met. See pages 1-2, lines 34 and 1 respectively, of the Specification.

23. In regards to claim 4:

Method according to claim 3, characterized by the fact that, for each position to be assumed by the model in said synthesis phase, the transforms are applied only to the vertices of the wire-frame corresponding to the feature points and the transforms are extended to the remaining vertices by means of a convex combination of the transforms applied to the vertices of the wire-frame corresponding to the feature points.

24. Ostermann discloses:

- If a FAP like a smile causes flexible deformations of the face model, the animation results in updating vertex positions of the affected IndexedFaceSet nodes. The affected vertices move along piece-wise linear trajectories that approximate flexible deformations of a face. A vertex moves along this trajectory as the amplitude of the FAP varies. The FaceDefTable defines for each affected vertex its own piece-wise linear trajectory by specifying intervals of the FAP amplitude and 3D displacements for each interval. See page 3, column 2 paragraph 4.

Claim Rejections - 35 USC § 103

25. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

26. Claims 2 and 5 are rejected under 35 U.S.C. 103(a) as being unpatentable over Ostermann (Animation of Synthetic Faces in MPEG-4), as applied to claims 1, 3 and 4.

27. In regards to claim 2:

Method according to claim 1, characterized by the fact that the coefficients of said convex combinations are functions of the following type:

$$\beta_n(t)=\begin{cases}\cos^2\left(\frac{\pi}{2}\frac{t-t_n}{t_{n+1}-t_n}\right);t\in[t_n,t_{n+1}]\\ \cos^2\left(\frac{\pi}{2}\frac{t-t_n}{t_{n+1}-t_n}\right);t\in[t_{n-1},t_n]\\ 0;t\in[t_{n-1},t_{n+1}]\end{cases}$$

28. Ostermann discloses:

- The rationale of the rejection of limitation (c) of claim 1 is incorporated herein.

29. Ostermann fails to explicitly disclose the three functions recited above in a purely mathematical form. However, the properties of said functions and their intended purpose are recited by Ostermann and thus are considered equivalent.

30. It would have been well to known and obvious to one skilled in the art, at the time of the applicant's invention, to conclude that for the three formulas disclosed above the result of the computations for $\beta_n(t)$ would yield a value of 0, for all values of *t*, due to the closed interval declaration of $t\in[t_{n-1},t_{n+1}]$.

31. In regards to claim 5:

Method according to claim 1, characterized by the fact that said visemes are converted into co-ordinates of the feature points of the face of the speaker, followed by conversion of said co-ordinates into said low-level facial animation parameters, as described in standard ISO/IEC 14496.

32. Ostermann discloses:

- The rationale of the rejection of limitation (a) of claim 1 and the rejection of claims 3-4 are incorporated herein.

33. It is noted that a viseme is considered part of the FAP set (68 parameters, consisting of both high and low-level parameters). FAPs are tied to the feature points of a given model. Thus, FAPs can be defined by the manipulation of said feature points of a given model. In addition for any high-level parameter, such as a viseme, there also exist low-level parameters, which when used in combination form elements of this high-level parameter, all of which fall under the title of FAP.

34. It would have been well known and obvious to one skilled in the art, at the time of the applicant's invention, that for any animation using a viseme, defined as a high-level FAP parameter, the relevant elements (feature points and high/low-level FAPs) which are embodied in said viseme are in turn processed according to their hierarchical layout (i.e. by vertex layout) in a given model. Because FAPs and feature points are so closely tied together one cannot modify one without somehow modifying the other in turn. As a result to modify one requires some modification of the other, regardless of the order in which this is accomplished.

35. Claims 6-7 are rejected under 35 U.S.C. 103(a) as being unpatentable over Ostermann (Animation of Synthetic Faces in MPEG-4), as applied to claims 1-5, in view of Guenter et. al (U.S. Patent No. 6, 072, 496).

36. In regards to claim 6:

Method according to claim 5, characterized by the fact that said low-level facial animation parameters, representing the co-ordinates of feature points, are obtained by analyzing the movement of a set of markers which identify the feature points themselves.

37. Guenter et. al discloses:

- The method uses the markers to generate a set of 3D points that act as control points to warp the base mesh of the 3D object being modeled. See column 6, lines 46-48.

38. It is noted that control points are considered equivalent to feature points in terms of how they behave.

39. In regards to claim 7:

(a) a sub-set of markers are associated to a stiff object applied to the forehead of the speaker;

40. Ostermann fails to explicitly disclose a sub-set of markers associated to a stiff object applied to the forehead of the speaker.

41. Guenter et. al discloses:

- Markers are applied to the actor's face. Then, multiple cameras at positions around the actor's face simultaneously record video sequences of the actor's face as the actor talk and emotes. See column 5, lines 14-17.
- The markers used in the test case were 1/8" circular pieces of fluorescent colored paper in six different colors. See column 6, lines 14-16.
- While fluorescent paper dots were used in the test case, a variety of different types and colors of markers can be used. See column 6, lines 39-40.

42. It would have been well known to one skilled in the art, at the time of the applicant's invention, to use one of a variety of conventional motion capture systems, which rely on the use of markers, tracked by a recording device, for the recording of

various body movement – specifically that of facial movement. Thus, it would have been obvious to apply markers to various locations on the face of a subject (i.e. the forehead) so to allow the recording the facial movements, which may be used for creating visemes and/or tracking facial changes.

(b) the face of the speaker is set, a the beginning of the recording, to assume a position corresponding as far as possible to the position of a neutral face model, as defined in standard ISO/IEC 14496, and a first frame of the face is such neutral position is obtained;

43. Ostermann discloses:

- MPEG-4 allows the encoder to completely specify the face model the decoder has to animate. This involves defining the static geometry of the face model in its neutral state. See page 2, column 2 paragraph 2.

44. Ostermann fails to explicitly disclose a first frame of the face is such neutral position is obtained.

45. Guenter et. al discloses:

- To compute the transform for our test case, we took a frame in which the subject had a neutral face. See column 11, lines 56-58.

46. It would have been well known to one skilled in the art, at the time of the applicant's invention, to record an initial first frame (reference frame) of a marked subject, who is to be recorded with the aid of markers, in a neutral state before any animation/movement is to occur so to supply reference values for later calculations after movement is made. Thus, it would have been obvious to combine this conventional

method of establishing a reference frame with that of a system for tracking movement over both time and space for any application in which movement is to be recorded for later processing, be it with or without audio elements.

(c) for all frames subsequent to the first frame, the sets of co-ordinates are rotated and translated so that the co-ordinates corresponding to the markers of said sub-set coincide with the co-ordinates of the markers of the same sub-set in the first frame.

47. Ostermann fails to explicitly disclose for all frames subsequent to the first frame, the sets of co-ordinates are rotated and translated so that the co-ordinates corresponding to the markers of said sub-set coincide with the co-ordinates of the markers of the same sub-set in the first frame.

48. Guenter et. al discloses:

- The process of labeling markers begins by first locating (for each camera view) connected components of pixels that correspond to the markers. The 2D location for each marker is computed by finding the two dimensional centroid of these connected components. The labeling process then proceeds by computing a correspondence between 2D dots in different camera views. Using the 2D locations in each camera view, the labeling process can then reconstruct potential 3D locations of dots by triangulation. To track motion of the markers over time, the labeling process starts with a reference set of dots based on the dot locations in the initial base mesh and pairs up this reference set with the 3D

locations in each frame. This gives a unique labeling for the dots that is maintained throughout the video sequence. See column 6, lines 47-67.

49. It would have been well to know to one skilled in the art, at the time of the applicant's invention, to establish a reference position for markers, on a given subject who is to be recorded, by which the movements (rotation, translation, etc.) of the subject are recorded through tracking these markers through both time and space. Thus, it would have been obvious to incorporate this conventional method of motion capture for the use in a system in which the movement of a subject's face, over both time and space, are to be recorded for possible later processing.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Peter-Anthony Pappas whose telephone number is 703-305-8984. The examiner can normally be reached on M-F 9:00am-6:30pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Mark Zimmerman can be reached on 703-305-9798. The fax phone number for the organization where this application or proceeding is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is 703-305-3900.

Peter-Anthony Pappas
Examiner
Art Unit 2671

PAP

MARK ZIMMERMAN
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2600